Computational and Mathematical Organization Theory
https://doi.org/10.1007/s10588-019-09297-2

S.I. : SOCIAL CYBER-SECURITY

To illuminate and motivate: a fuzzy-trace model of the spread of information online

David A. Broniatowski[1] · **Valerie F. Reyna**[2]

Abstract
We propose, and test, a model of online media platform users' decisions to act on, and share, received information. Specifically, we focus on how *mental representations* of message content drive its spread. Our model is based on fuzzy-trace theory (FTT), a leading theory of decision under risk. Per FTT, online content is mentally represented in two ways: verbatim (objective, but decontextualized, facts), and gist (subjective, but meaningful, interpretation). Although encoded in parallel, gist tends to drive behaviors more strongly than verbatim representations for most individuals. Our model uses factors derived from FTT to make predictions regarding which content is more likely to be shared, namely: (a) different levels of mental representation, (b) the motivational content of a message, (c) difficulty of information processing (e.g., the ease with which a given message may be comprehended and, therefore, its gist extracted), and (d) social values.

Keywords Gist · Verbatim · Vaccines · Misinformation · twitter

1 Introduction

In this paper, we present, and test, a novel mathematical formulation of how information spreads online. Our model is based on fuzzy trace theory (FTT)—a leading account of decision under risk—which emphasizes the combined roles of mental representation, message content, social values, and individual differences. FTT posits that individuals encode multiple representations of a stimulus, such as online information, in parallel. These representations are referred to as *gist*—the essential

✉ David A. Broniatowski
broniatowski@gwu.edu

1 Department of Engineering Management and Systems Engineering, School of Engineering and Applied Science, The George Washington University, 800 22nd St. NW #2700, Washington, DC 20052, USA

2 Human Neuroscience Institute, Center for Behavioral Economics and Decision Research, and Cornell Magnetic Resonance Image Facility, Cornell University, Ithaca, USA

Published online: 12 August 2019

meaning of the information—and *verbatim*—a detailed symbolic representation of the stimulus. Thus, our model incorporates factors capturing the extent to which the individual "gets the gist" (i.e., is able to extract meaningful information) and is therefore *illuminated* by the content of the message. We also incorporate a factor capturing *motivation* to share, such as may be triggered by material containing especially compelling media (e.g., vivid photos or surprising content). Our formulation is novel because it combines these cognitive and motivational considerations into a common computational framework.

1.1 Rationale for our approach

Our approach is motivated by the widespread effects of online misinformation and disinformation across multiple contexts. Although we focus on misinformation about vaccines in this article, online misinformation is now widely considered to be a threat in multiple domains (Grinberg et al. 2019), underlying shifts in electoral politics (Swire et al. 2017), epidemic outbreaks (Chou et al. 2018), and several other areas pertinent to national security. Indeed, social media have an especially wide reach, with Twitter as one of the most popular platforms. A recent Pew Center study (Perrin 2015) indicates that more people get their news from social media than from any other source. As of March 1, 2018, 24% of all online adults, and 45% of adults aged 18–24, are on Twitter (Smith and Anderson 2018). Thus, social media, and especially Twitter, enable the rapid increase in the speed and scope of dissemination of narratives that may affect decisions and other behaviors, including decisions to share information online.

Public health professionals face challenges from this new communications environment (Chou et al. 2018). For example, the World Health Organization has recently declared vaccine hesitancy to be one of the world's top 10 public health threats[1]—in large part, driven by anti-vaccine sentiment on social media (Brewer et al. 2017). Indeed, the journal *Vaccine* devoted an entire special issue to the role social media plays in vaccination decisions (Betsch et al. 2012). Importantly, the consensus article in this special issue emphasizes the role of psychological factors, such as how online narratives are processed, in the spread of online information. Specifically, they state that "Narratives have inherent advantages over other communication formats...[and] include all of the key elements of memorable messages: They are easy to understand, concrete, credible...and highly emotional. These qualities make this type of information compelling..." (p. 3730). Furthermore, within the specific domain of vaccine refusal, recent studies have documented the role of both domestic and state-sponsored foreign actors using misinformative online messages about public health topics to market products and to promote political discord (Broniatowski et al. 2018; Jamison et al. 2019; Subrahmanian et al. 2016).

[1] Ten threats to global health in 2019. https://www.who.int/emergencies/ten-threats-to-global-health-in-2019. Accessed 14 Mar 2019.

Although our primary focus in this paper is misinformation about vaccines, recent political developments have highlighted the popularity of "fake news" which, although factually inaccurate, may have been shared more widely online than vetted media sources (Silverman 2016; and see also Dredze et al. 2017). Findings suggest that unverified information with highly surprising, or emotionally arousing and therefore motivational, content may travel faster and farther than information containing verbatim facts (Vosoughi et al. 2018; Berger and Milkman 2012). We propose that the *influence* of news, and its concomitant recognition as fake or genuine, can be studied as a scientific problem (see also Pennycook et al. 2018). We therefore focus on analogues in the literature (i.e., vaccination) that provide theoretical and empirical insight into the process of influence through social media.

In this paper, we aim to model how these psychological factors drive online sharing. Studies in psycholinguistics have identified a narrative's "coherence" as a key factor driving a story's comprehensibility and long-term retention (Trabasso et al. 1982; Van den Broek 2010; Pennington and Hastie 1991). Although several dimensions of narrative coherence have been proposed (Reese et al. 2011; Gernsbacher et al. 1996), there is a consensus in the literature that coherent narratives often provide a *causal* structure for the events described (Mandler 1983; Trabasso and Sperry 1985; Gernsbacher et al. 1990; Diehl et al. 2006; Van den Broek 2010), therefore conveying the meaning, or gist of the story. In contrast, incoherent stories contain a relatively weak causal structure. According to this reasoning, therefore, online information facilitating causal coherence produces more coherent and meaningful gists, and will therefore be more influential. In contrast, official communications tend to focus on literal verbatim facts without emphasizing the causal relations among those facts in a manner that communicates a coherent gist. For example, government sites tend to focus on "how" vaccines work, whereas anti-vaccination narratives focus on providing a causal (though not necessarily accurate) explanation for "why" vaccines are harmful and are consequently more comprehensible, influential, and memorable (Trope and Liberman 2010; Fukukura et al. 2013).

The outline of this paper is as follows: In Sect. 2, we provide an overview of literature motivating the use of FTT to model the spread of online information. In Sect. 3, we provide a description of our modeling approach. Section 4 tests this model on an existing dataset of tweets about vaccines. Finally, Sect. 5, discusses the implications of these findings for future work, and concludes.

2 Literature review

2.1 Fuzzy-trace theory

According to FTT, effective messages help readers retain the meaning of the message in memory (because gist endures) and, hence, facilitate availability of the knowledge at the time of behavior. FTT can be used to explain the popularity of online messages because of the search for meaning and the tendency to interpret events even when knowledge is inadequate. FTT's approach to online communication builds on the core concepts of gist and verbatim mental representations, modified and adapted from the psycholinguistic literature (Kintsch 1974) but modified in the light of more recent findings (see Reyna 2012). According to FTT, meaningful stimuli such as social media messages (e.g., those that communicate narratives) are encoded into memory in two forms: a verbatim representation (the objective stimulus or a decontextualized representation of what actually happened) and a gist representation, the subjective or meaningful interpretation of what happened (Reyna et al. 2016). Verbatim representations encode details, such as exact numbers. For example, an anti-vaccine message discussing the results of an isolated scientific study (Cowling et al. 2012) out of the context of the broader literature (Sundaram et al. 2013) may state that "Flu Shot Induces 4.4-fold increase in non-flu acute respiratory infections." In contrast, a gist representation encodes the essential meaning of the sentence. Furthermore, there may be multiple gist representations. An uninformed gist that supports avoiding the vaccine might be held by a non-expert as follows: "Say no to the Flu Shot !! It's ineffective and dangerous ..." In contrast, a gist held by an expert might emphasize "many problems w/ reporting bias & confounding" indicating that the findings of this specific study should not be considered definitive. Gist representations depend on culture, knowledge, beliefs, and other life experiences (Reyna and Adam 2003). However, in practice, coherent gist representations have been communicated to diverse audiences. Importantly, gist interpretations, rather than verbatim facts, tend to guide decisions and behavior.

When making sense of text, gist representations reflect coherent, causal stories (Reyna 2012; Reyna et al. 2016; Pennington and Hastie 1991). These narratives "connect the dots," to offer a coherent account, and are more likely to be accepted. More coherent stories such as those connecting adverse health outcomes (e.g., autism) to certain behaviors (e.g., vaccination), are more likely to be accepted because they "make sense"—i.e., they provide an explanation for otherwise mysterious adverse events. Online messages are predicted to increase in popularity when similar messages from one's friends or other trusted sources, make certain ideas plausible (e.g., that the government would intentionally infect people), especially coupled with an increased prevalence of poorly understood outcomes. Thus, a story describing how children developed symptoms of autism after having gotten vaccinated might allow one to erroneously conclude that vaccines cause autism. (In fact, the symptoms of autism tend to occur around the same time as the US Centers for Disease Control and Prevention recommend that children receive vaccines.) Similar spurious correlations underlie the false claims that exposure to the larvicide pyriproxifen (Vazquez 2016) or receipt of the DTaP vaccine by pregnant mothers, rather than the Zika virus, causes birth defects (Dredze et al. 2016a).

2.1.1 Individual differences

Prior work on FTT has shown that an individual's reliance on gist vs. verbatim representations is associated with individual differences in metacognitive monitoring and editing initial reactions to information (Broniatowski and Reyna 2018). In the domain of risky decision problems that involve numerical information, studies have found that more *numerate* individuals (i.e., those possessing greater mathematical ability) are less prone to framing biases (Liberali et al. 2012; Peters et al. 2006;

Peters and Levin 2008; Schley and Peters 2014), suggesting an increased ability to directly compare decision options that have the same verbatim expected value (Broniatowski and Reyna 2018). (Framing biases are effects of phrasing the same outcomes differently, such as phrasing choice options in terms of saving 200 people or as 400 people dying when 600 people are expected to die if nothing is done). Similarly, subjects exhibiting high *Need for Cognition* (NFC) (Cacioppo et al. 1984; Cacioppo et al. 1996) tend to be more consistent across multiple exposures to framing problems, presumably because they are able to identify the common structure of these problems (Broniatowski and Reyna 2018; LeBoeuf and Shafir 2003; Simon et al. 2004; Curseu 2006). However, the effects of numeracy and NFC do not explain where framing biases come from to begin with—namely, from gist representations of meaning in context—but they do explain the tendency to inhibit gist, especially in within-subjects designs featuring different frames for the same information (LeBoeuf and Shafir 2003; Broniatowski and Reyna 2018).

Individual differences have been found in the domain of narrative comprehension (Rapp et al. 2007). For example, Linderholm et al. (2000) and Van den Broek (2010) found that more skilled readers, and those with more relevant background knowledge, were better able to extract the gist from narratives with poorly-defined causal structures. In addition, LaTour et al. (2014) observed that subjects higher in NFC were better able to identify and reject narratives whose gists were inconsistent [see also Pennycook and Rand (2018) who found that subjects scoring higher on the cognitive reflection test (Frederick 2005) were better able to distinguish between true and misinformative headlines—cognitive reflection is known to be correlated with both numeracy (Liberali et al. 2012; Cokely and Kelley 2009) and need for cognition (Pennycook et al. 2016)]. More recently, van den Broek and Helder (2017) describes evidence for a model of narrative comprehension in which multiple levels of mental representation are encoded. Specifically, the authors differentiate between readers who prefer to use coherence-building strategies relying on effortful "close-to-the-text" reading (perhaps analogous to those exhibiting high NFC) and those who utilize a more interpretive strategy that is "farther" from the text. Importantly, such interpretive processes are associated with domain expertise (Goldman et al. 2015)—a hallmark of gist processing (Reyna and Lloyd 2006). Thus, there is reason to believe that individual differences associated with systematic variation in susceptibility to framing biases may also be associated with differences in one's ability to extract a meaningful gist from online narrative text. Furthermore, a subject's ability to extract this meaningful gist is a function both of the subject's characteristics and the narrative's content—more difficult texts are likely to appeal only to those subjects possessing the willingness and ability to expend the effort to comprehend them.

2.1.2 Motivational factors

Beyond the effects of metacognitive monitoring and editing, there is evidence indicating the role of motivational factors in risky decisions. For example, reward sensitivity has been associated with risk-taking across a wide range of problem types (e.g., Reyna et al. 2011; Broniatowski and Reyna 2018; Galván 2017). Berger and

Milkman (2012) examined the psychological drivers of online information diffusion with implications for the motivational factors posited by our model. Specifically, the authors examined the determinants of what makes specific news articles more likely to be shared by email. They found that "virality" can essentially be described by two classes of factors: (1) motivational factors, which they described as "how surprising, interesting, or practically useful content is (all of which are positively linked to virality)..."; and (2) emotional valence/arousal factors. Regarding the latter, positively valenced items were more likely to be shared than negatively valenced items; however, arousal plays an important role as well. Specifically, high-arousal emotions, such as awe, anger, or anxiety were more likely to be viral whereas low-arousal emotions, such as sadness lead to less virality. In the domain of online narrative, such factors may also include flashy media, surprising or otherwise emotionally-arousing content (Vosoughi et al. 2018; Berger and Milkman 2012) and other motivational "clickbait" designed to temporarily grab the user's attention. Additionally, motivational factors include trust in specific sources including the government, celebrities, and other opinion leaders (e.g., Quinn et al. 2013; Swire et al. 2017), and prior associations that trigger strong impulsive reactions (e.g., appeals to emotion). Although these factors typically engender virality, their effects tend to quickly diminish (Swire et al. 2017).

2.2 Evidence for FTT's predictions online

2.2.1 Explicit tests of FTT online

Prior work (Broniatowski et al. 2016) has examined FTT's predictions in the context of the Disneyland Measles Outbreak which began in December 2014 at Disneyland in California and led to 111 confirmed cases of measles in seven states (as well as in Canada and Mexico). Although measles was widely considered eliminated in the United States, reduced vaccination rates in some communities, due to concerns about vaccine toxicity, ultimately called attention to the issue of herd immunity—how slight reductions in vaccination rates can lead to epidemics.

This study was conducted in the context of an ongoing debate: Does including an anecdotal narrative lead to more effective communication compared to presenting "just the facts" (Buttenheim and Asch 2016) (i.e., statistical data)? In addition to the perceived effectiveness of narratives noted above, public health officials have been hesitant to include stories in their communications due to concerns of appearing biased or paternalistic. In contrast, FTT predicts that the verbatim details of a message are incorporated separately from, but in parallel to, the gist of the message. According to FTT, narratives are effective to the extent that they communicate a gist representation of information that then better cues motivationally relevant moral and social principles.

Broniatowski et al. (2016) crowdsourced the coding of 4581 out of a collection of 39,351 outbreak-related articles published from November 2014 to March 2015, asking coders to indicate whether each article expressed statistics (a verbatim representation), a story, and/or a "bottom line meaning" (i.e., a gist). Finally, they measured how frequently these articles were shared on Facebook. Results were consistent with expectations based on FTT, enumerated below:

1. FTT predicts that gist and verbatim representations are encoded in parallel. The authors found that both gist and verbatim types of information were associated with an article's likelihood of being shared at least once, constituting distinct sources of variance.
2. The effects of gist were larger than the effects of verbatim, consistent with FTT's "fuzzy-processing preference."
3. Stories did *not* have a significant impact on an article's likelihood of being shared after controlling for gist and verbatim, indicating that stories are only effective to the extent that they communicate a gist.
4. Among those articles that were shared at least once, only the expression of a gist was significantly associated with an increased number of Facebook shares (articles with gists were shared 2.4 times more often, on average, than articles without gists).
5. Articles expressing a gist that also expressed positive opinions about both pro- and anti-vaccine advocates were shared 57.8 times more often than other articles, suggesting that facts can indeed be effectively shared if concerns of those on the "opposing" side are acknowledged (while emphasizing the bottom-line meaning of the data in its cultural context).
6. Motivational factors—e.g., presence of vivid media—were associated with an article's likelihood of being shared at least once, but not with more than one share.

These results provide evidence supporting FTT's expectations for online information sharing and suggest that content features should be predictive of the spread of online misinformation. Furthermore, there is evidence supporting the combined roles of meaning-making (gist) and motivation on the sharing of online information. As will be discussed below, our model incorporates both types of factors.

2.3 Content features have not traditionally incorporated gist

Although online misinformation and disinformation are relatively new problems, significant work has been performed examining the spread of ideas, such as rumors, through social networks (e.g., Rogers 2010). Most of this prior work has focused on complex contagion (Centola 2010; Mnøsted et al. 2017) and homophily (Centola 2011; Bakshy et al. 2015; Grinberg et al. 2019)—both mediated by social network structure—as antecedents of information sharing. In contrast, comparatively little analysis of the psychological content of this information has been performed.

Romero et al. (2011) conducted an observational study that was explicitly designed to examine the role of content while controlling for social network factors. Specifically, they defined two content-based measures of a twitter hashtag's spread: (1) "stickiness", probability of sharing based on at least one exposure to a hashtag and (2) "persistence"—whether sharing will continue to occur after multiple exposures. Using this approach, the authors found evidence in support of variation in complex contagion by topic. Although this analysis primarily focused on hashtags rather than on true semantic content, the authors did provide some evidence suggesting that more *meaningful* hashtags, indexing topics such as politics, may be more persistent over time when compared to less meaningful "idiomatic" hashtags that are more motivational in nature.

In general, several studies have focused on verbatim-level features, rather than the gist-based semantic content that FTT predicts would be compelling. These studies have concluded that verbatim content features are not predictive when compared to structural features. For example, Petrovic et al. (2011) used a machine-learning approach to examine the relative predictive power of "social features" (features of the tweet's author) compared to "tweet features" (text and statistical verbatim features of the tweet) in predicting retweets. The authors found that social features, rather than verbatim content features, were most predictive of retweets. Similarly, Cheng et al. (2014) found that the size of an information cascade could be more easily predicted based on temporal (i.e., how quickly an item was shared after having been initially posted) and structural, rather than verbatim content-based, features as the cascade grew. Tsur and Rappoport (2012) concluded that structural features of twitter hashtags captured more variance than did verbatim content features. However, unlike prior studies, Tsur and Rappoport (2012) examined the *context* of tweets, finding that an interaction of content and contextual features were indeed predictive of a hashtag's spread, adding significant predictive value above the contribution of structural features. The authors acknowledge that the cognitive/psychological attributes of their tweets were not well characterized, potentially explaining why they were unable to capture more variance with these factors. There is therefore a need to explicitly examine the role of gist factors in the spread of information online.

3 A model of information sharing online

We aim to explicitly test FTT's core constructs using social media data. Our approach builds on a recent mathematical formalization of FTT (Broniatowski and Reyna 2018).

3.1 Parameter specification

The structure of the model is as follows: FTT posits a hierarchy of gist that is, in the domain of numbers, analogous to scales of measurement (Reyna and Brainerd 2008; Stevens et al. 1946). Broniatowski and Reyna (2018) illustrates this hierarchy with the following example:

> ...consider the following choice between:
>
> 1. Winning $180 for sure; versus
> 2. 0.90 chance of winning $250 and 0.10 chance of no money.

[At the simplest level of gist], people represent this decision as a categorical choice between the following two options:

1. Some chance of winning some money
2. Some chance of winning some money

Given this representation, most decision makers would favor option 1 because it promises some money without the chance of no money. However, more precise, yet still qualitative, representations are also generated simultaneously, such as ordinal representations (e.g., small vs. large amount of money):

1. More chance of winning less money
2. Less chance of winning more money and some chance of winning no money.

This representation does not allow for a clear decision to be made because most people would prefer winning more money to winning less money, but they would also prefer more chance of winning to less chance of winning. Finally, one may choose a precise interval representation of the problem whereby one calculates the expected value of each option by multiplying its respective outcomes by their probabilities, as follows:

1. Expected value of \$180 (i.e., \$180 × 1)
2. Expected value of \$225 (i.e., \$250 × 0.90 + \$0 × 0.10)

Given this representation, most decision makers would favor option 2.

According to FTT, in the domain of numbers, risky decisions are encoded at the categorical, ordinal, and interval levels simultaneously. Broniatowski and Reyna (2018) models the probability, P, that a subject will choose a given decision option in a risky choice gamble by the logistic function,

$$P(\mathbf{x}) = \frac{1}{1 + e^{-(\mathbf{a}\cdot\mathbf{x}+b)}} \tag{1}$$

where $\mathbf{x}$ is a vector containing an entry for each level of mental representation (e.g., gist and verbatim), and $\mathbf{a}$ is a vector containing an entry corresponding to decision weights that are associated with individual differences in a subject's ability to inhibit cognitive biases e.g., due to a subject's numeracy (e.g., Liberali et al. 2012; Peters et al. 2006; Peters and Levin 2008; Schley and Peters 2014) and NFC (Cacioppo et al. 1984; Cacioppo et al. 1996). In addition, b, captures a subject's overall motivation. We account for conflict between representations by adding weighted votes from each representation. As in the example above, preferences depend on the application of social values (e.g., winning money is good) to representations of options. If this gist representation prefers the certain option (− 1), the ordinal representation is indifferent, and the expected value representation of the problem prefers the risky option (+ 1), and then $\mathbf{x} = [-1\ 0, +1]$. Furthermore, research suggests that

Table 1 Summary of model parameters

Parameter	Explanation
$\mathbf{x}$	Mental representation of prior knowledge, consisting of a vector of elements with values -1, 0, or 1
$\mathbf{a}$	Weights assigned to each mental representation, consisting of a vector of weights for each element in $\mathbf{x}$
b	Scalar parameter capturing strength of motivational factors

neurotypical adults weigh the simplest gist representation most heavily; for example, if the representational weights $\mathbf{a} = [2,1,1]$ for each of the levels of mental representation posited by our model, then

$$\mathbf{a} \cdot \mathbf{x} = -2 + 0 + 1 = -1 \tag{2}$$

Finally, suppose we estimate $b = 0.5$, indicating a preference for the more rewarding, though riskier, option ($b > 0$ indicates overall risk-seeking behavior, whereas $b < 0$ indicates risk averse behavior). Under these assumptions, the probability that a randomly chosen subject from our sample will choose the risky gamble option is

$$P(\mathbf{x}) = \frac{1}{(1 + e^{-(-1+0.5)})} = 38\% \tag{3}$$

Inputs to this model are the three parameters outlined above ($\mathbf{x}, \mathbf{a}$, and b) and the model outputs a prediction regarding a decision probability (summarized in Table 1).

3.1.1 Mental representation of prior knowledge

Many online audiences lack extensive prior knowledge about controversial topics. Under these circumstances, causal narratives that provide explanations for otherwise mysterious adverse events are easy to comprehend and therefore compelling. Under such circumstances, individuals also rely on their social contacts for signals of the trustworthiness of online information. For example, Granovetter and Soong (1983) described the decision to adopt a behavior, such as spreading a rumor, as a function of the number of friends who had done the same. Specifically, they framed this decision as a risky binary choice: sharing when few people have done so is risky, yet doing so when many others have done so is safer. Thus, we posit that the perception of whether sharing a controversial article is perceived as risky is affected by social influence as determined by a threshold. When the number of exposures does not exceed the threshold, sharing the article is perceived as "risky": here, the decision-maker faces the following binary choice analogous to a framing problem (Tversky and Kahneman 1981) (see also Broniatowski et al. 2015; Klein et al. 2017):

A. Don't share the online information and lose no social capital
B. Share the article and maybe lose no social capital, but maybe lose some social capital (such as when one is criticized by friends)

In contrast, when the number of exposures exceeds the threshold (meaning that the information is now socially validated), the decision-maker faces the following choice:

C. Don't share the article and gain no social capital for sure
D. Share the article and maybe gain some social capital as reflected by likes, reshares, etc., but maybe gain no social capital

The factor of representations is captured by the **x** vector in our model.

3.1.2 Values associated with gist principles

To decide between the options outlined above (A vs. B or C vs. D), subjects must apply social values that they endorse, called "gist principles" in FTT because, like information, they are mentally represented in simple gist forms. For example, a subject who is seeking social approval and who perceives the possibility of positive attention from sharing will be more likely to choose to share the information. They apply the gist principle—that "positive attention is better than no attention"—to their representations of options. Similarly, one who feels that he or she is at risk of social opprobrium but perceives nil risks from not sharing would not share the information since not getting criticized is preferred to getting criticized. This factor is captured by the signs of the elements in the **x** vector. Naturally, the subject's assessments of how their friends might react are central to their judgments, with culture, worldview, and social identity all informing the gist of what information will be well received when shared.

3.1.3 Weights assigned to each mental representation

Subjects differ in the degree to which they rely on categorical gist (and other gist representations) versus literal verbatim information. For example, those who are more numerate (in the sense of rote computation) can rely more on precise numerical and literal details, giving less relative weight to categorical gist, all else equal (Reyna and Brainerd 2008). Similarly, when cued with obvious equivalencies, such as when framing is manipulated within-subjects, those with higher Need for Cognition may compare between frames, giving more relative weight to verbatim tradeoffs. Individual differences in reliance on these representations are captured by the **a** vector. Conversely, social media content that is easier to comprehend, because it is less detailed, is likely to be more widely shared.

3.1.4 Motivational factors

Motivation and strong emotion can bias decisions. For example, articles may contain "clickbait" or other factors that are designed to trigger impulsive sharing behavior. This factor is captured by the b parameter.

Granovetter and Soong (1983, p. 167) presage these factors in their threshold model of collective action. Specifically, they associate risky decisions with motivational personality factors (“Some individuals are more daring than others”), factors associated with social values in cultural context (“some are more committed to radical causes...”), and factors associated with verbatim cost-benefit calculations (“rational economic motives”).

3.2 Dataset

In order to test our model, we must measure its key constructs on social media. The analysis that follows is based on a set of 10,000 tweets about vaccines collected between November, 2014 and September, 2017, tagged as relevant to vaccines using the classifier described in Dredze et al. (2016b), and containing at least one word starting with “vax” or “vacc”.[2] This procedure yielded a dataset that was largely relevant to the online discourse about vaccine safety, although with some outliers (such as tweets pertaining to vaccinating pets and messages from fans of a band called “The Vaccines”). We chose not to remove these tweets since they were segmented by the topic model analysis (described below). Each tweet was hand-annotated by three raters as pro-vaccine, anti-vaccine, or neutral. Annotators had moderate agreement (Fleiss’ $\kappa = 0.49$) in the first round of annotation, and annotation rounds were conducted until raters reached consensus (typically 2–3 rounds; see Broniatowski et al. (2018) for full dataset details, annotation instructions, and procedure).

3.3 Operationalizing model parameters

The factors identified above map to the key elements of the model of decision under risk upon which we build (Broniatowski and Reyna 2018). We assume that the probability that a given individual will share an item of information can be described using the logistic function described in Eq. (1) where $\mathbf{x}$ is a vector capturing mental representations, $\mathbf{a}$ is a vector capturing weights placed on each such representation, and b is a scalar capturing motivational factors.

3.3.1 P(x)—the probability that a given message is shared

Our model may be rewritten as

$$logit[P(\mathbf{x})] = \mathbf{a} \cdot \mathbf{x} + b \tag{4}$$

where

$$logit[P(\mathbf{x})] = log\left(\frac{P(\mathbf{x})}{1 - P(\mathbf{x})}\right) \tag{5}$$

We operationalize P($\mathbf{x}$) by measuring the total number of retweets per follower for each message in our dataset. We use a logistic transform so that we may test our predictions using linear regression models.

[2] Corresponding tweet ids may be found at https://github.com/broniatowski/Illuminate-and-Motivate.

3.3.2 x—mental representation

Although we cannot directly measure the mental representations of every social media user in our sample, we may examine proxies for gist. Specifically, Griffiths et al. (2007) posited that Latent Dirichlet Allocation (LDA) (Blei et al. 2003) may be used as a measure of the gist associated with a given document. Although we agree that probabilistic topics, such as those generated by LDA, may be associated with gist, there is work demonstrating that LDA does not always yield topics that are comprehensible by humans (Chang et al. 2009)—some topics are expected to be more coherent than others. As indicated above, we expect that topics that "connect the dots"—i.e., expressing causal coherence—are more likely to capture a compelling gist.

In order to capture a proxy for gist, we fit a 50-topic LDA model to our dataset using unigram and bigram features using the scikit-Learn (Pedregosa et al. 2011) and lda (Riddell 2014) python packages. This allows us to determine the probability that any given tweet is about a given topic. These probabilities were converted into logarithmic units using a logistic transform to control for floor and ceiling effects. The top five most frequent terms associated with each topic are shown in Table 2.

Topic 2, in particular, captures the gist that vaccines cause autism. Since this topic explicitly captures a causal gist, we expect that it will be associated with a higher number of retweets per follower.

3.3.3 a—representational weights

Although our data do not allow us to directly measure the weights that each sharer places on different mental representation, we are able to determine the comprehensibility of each tweet using standard metrics. We expect that tweets that are more difficult to comprehend will be shared less frequently because some individuals, perhaps those with lower literacy, will not have the ability to derive meaningful information from them, whereas those with higher Need for Cognition may understand them but may prefer to share something more compelling.

We assess the comprehensibility of a tweet using several standard measures contained in the textstat python package (Bansal 2018). Since several of these measures are correlated, we conducted a principal component analysis (PCA) to extract orthogonal factors associated with text comprehensibility, corresponding to readability, verbatim features, and number of sentences (see Table 3). These principal components were used as predictors.

3.3.4 b—motivational factors

Many online messages contain compelling multimedia presentations, such as vivid images, movies, or sounds that are expected to be motivational. For example, previous work indicates that the presence of images on social media increases the likelihood that the message will be shared at least once (Broniatowski et al. 2016; Chen

Table 2 Topics extracted from dataset using LDA

Topic ID	Top five terms				
1	people	like	think	really	better
2	vaccines	autism	cause	cause autism	vaccines cause
3	doctors	immunization	vaccineswork	cancer	causing
4	vaccine	cough	whooping	whooping cough	baby
5	brain	don	damage	shows	million
6	vaccine	report	influenza	drug	influenza vaccine
7	new	disease	prevent	help	new vaccine
8	polio	polio vaccine	gates	long	campaign
9	zika	zika vaccine	scientists	develop	won
10	vaccine	dr	youtube	video	team
11	kids	vaccinate	vaccinate kids	son	guinea
12	vaccines	pharma	big	years	death
13	vaccine	world	dengue	dengue vaccine	use
14	vaccine	hepatitis	time	used	hepatitis vaccine
15	children	mandatory	school	test	immunity
16	mmr	injury	mmr vaccine	vaccine injury	cases
17	vaccination	free	rabies	clinic	available
18	measles	work	good	outbreak	vaccine
19	vaccine	cdc	linked	autism	fda
20	Health	news	today	medical	public
21	malaria	malaria vaccine	dogs	come	gets
22	vaccinated	children	china	right	scandal
23	virus	zika	zika virus	know	need
24	hiv	new	hiv vaccine	africa	aids
25	hpv	hpv vaccine	women	young	girls
26	cancer	protect	cancer vaccine	say	years
27	vaccines	government	stop	debate	live
28	vaccines	autism	science	cdcwhistleblower	vaccines autism
29	vaccine	california	sb277	law	state
30	vaccines	gsk	developing	countries	pfizer
31	vaccine	meningitis	risk	dangerous	shingles
32	shot	need	year	tetanus	tetanus shot
33	amp	vaccinated	getting	tt	vaccines amp
34	vaccine	trials	human	clinical	protection
35	vaccine	spread	reuters	injured	cost
36	vaccine	make	sure	injuries	men
37	ebola	ebola vaccine	trial	market	global
38	vaccine	diseases	court	uk	caused
39	says	vaccinated	times	study	working
40	anti	news	save	gt	2015
41	vaccinations	dog	experts	rabies	end
42	flu	flu vaccine	flu vaccines	swine	swine flu

Table 2 (continued)

Topic ID	Top five terms				
43	effective	safe	safety	let	robert
44	vaccine	world	india	scientists	vrus vaccine
45	vaccine	research	know	did	does
46	vaccines	study	immune	researchers	testing
47	vaccination	fever	yellow	yellow fever	hpv vaccination
48	vaccinate	parents	children	vaccinate children	australia
49	given	polio	cancer	americans	virus
50	don	child	just	got	want

All word tokens have been converted to lower case and punctuation has been removed. For example, the token “don” refers to the word “don’t”

Table 3 Results of PCA applied to measures of text comprehensibility

	Readability (40%)	Verbatim (31%)	Sentences (17%)
Gunning-Fog Index	0.95		
Dale-Chall Readability Score	0.94		
Reading ease	– 0.91		
Flesch-Kincaid Index	0.88		
Automated Readability Index	0.8		
Difficult words	0.72	0.52	
Length		0.96	
Syllable count		0.95	
Lexicon count		0.93	
Linsear write formula		0.73	– 0.57
Sentence count			0.89
SMOG Index			0.81

Following Kaiser’s criterion (retaining all eigenvalues ≥ 1.0) three factors, explaining 88% of the variance in the data, were retained. Factor loadings ≥ 0.40 are shown

SMOG “simple measure of gobbledygook”

and Dredze 2018), presumably because it is more noticeable and often more emotionally arousing. Thus, we record whether a given tweet contains any such media as a proxy for its motivational power.

Additionally, we assess the emotional content of a tweet using two separate measures: (1) weighted emotion scores associated with Plutchik’s eight basic emotions (joy, trust, fear, surprise, sadness, anticipation, anger, and disgust) (Plutchik 2001)[3], and (2) mean valence, arousal, and dominance scores associated with the Affective

[3] Specifically, we added the raw weights from dictionaries of words (Mohammad and Turney 2013) and hashtags (Mohammad and Kiritchenko 2015) associated with these basic emotions.

Table 4 Results of PCA applied to measures of emotion

	ANEW (28%)	Negative (22%)	Positive (16%)
Valence	0.97		
Dominance	0.97		
Arousal	0.93		
Sadness		0.82	
Anger		0.76	
Disgust		0.70	
Fear		0.70	
Anticipation			0.74
Trust			0.65
Surprise			0.61
Joy	0.46		0.56

Following Kaiser's criterion (retaining all eigenvalues ≥ 1.0) three factors, explaining 65% of the variance in the data, were retained. Factor loadings ≥ 0.40 are shown

ANEW "Affective Norms for English Words" (Bradley and Lang 1999)

Norms for English Words (ANEW) dictionary (Bradley and Lang 1999). We again conducted a PCA to extract orthogonal factors associated with emotion, yielding three dimensions corresponding to ANEW scores (averaged across both positive and negative valences), Plutchik's negative emotions only, and Plutchik's positive emotions only (see Table 4). These principal components were used as predictors.

We tentatively associate these emotional measures with the *b* parameter because Vosoughi et al. (2018) and Berger and Milkman (2012) speculated that such emotions were the driving force behind virality (however, see Rivers et al. 2008 for a more extensive discussion of the relationship between different definitions of emotion and decision-making). Finally, we included a dummy variable indexing whether a tweet was generated by a "verified user"—defined by twitter as accounts "of public interest"[4]—and therefore a proxy for celebrity.

3.4 Regression analyses

The aim of our analysis is twofold. On one hand, fuzzy-trace theory provides an *explanation* for which tweets about vaccines are shared online. We therefore seek to determine which of the theoretically-motivated factors, identified above, are significantly associated with online sharing. On the other hand, we seek a model that may be used to *predict* which of these tweets are more likely to be shared on new data without overfitting (for more about the distinction between prediction and

[4] About verified accounts. https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts. Accessed 15 Mar 2019.

explanation, see Shmueli 2010). Our model selection process is informed by these two parallel, yet complementary, goals.

3.4.1 Data segmentation

Consistent with our prior work (Broniatowski et al. 2016), we separately analyzed the factors driving virality (i.e., the number of retweets per follower) from those associated with the likelihood that a given tweet was retweeted at least once. Consequently, after removing 2254 tweets that were generated by accounts with 0 followers (meaning that we could not calculate the number of retweets per follower), we separated our sample into two segments—those that had been retweeted at least once (n = 1388), and those that had not (n = 6358).

3.4.2 Linear multiple regression

Our goal was to select the best-fitting linear regression model to predict retweets per follower among those tweets that had been retweeted at least once. Consistent with our dual aims of prediction and explanation, we carried three rounds of model fitting, each of which used two separate model selection procedures.

1. *Explanation* We performed model fitting by bidirectional stepwise elimination, using the Akaike Information Criterion (AIC; Akaike 1976), which is mathematically equivalent to L_0-norm regularization) as the minimization criterion. The starting state for this stepwise procedure was a model including all main effects, but no interactions. Terms were removed or added one at a time if doing so reduced AIC.
2. *Prediction* We segmented our data into thirds, holding one segment out for measuring predictive accuracy, with the remaining tweets used for training and test. We used Least Absolute Shrinkage and Selection Operator (LASSO) regression—a technique based on L_1-norm regularization—with threefold cross validation to determine the factors underlying the most predictive model.

In both cases, predictors included items associated with theoretically-motivated factors:

1. $\mathbf{x}$: logit-transformed proportions of all 50 topics
2. $\mathbf{a}$: the three PCA dimensions of text comprehensibility
3. Tweet polarity: pro-vaccine, anti-vaccine, or neutral
4. b: the three PCA dimensions of emotion, dummy variables indicating the presence or absence of vivid media, and whether or not a tweet was generated by a verified user (an explicit measure of source credibility)

In addition, first- and second-order interaction terms between topics, polarity, and comprehensibility were included to account for their multiplicative effects in our model.

In the second round of model-fitting, we constructed new ordinary least squares (OLS) regression models containing only the factors that replicated across both the bidirectional elimination and LASSO model selection methodologies. Finally, in the third round, we removed all factors that were not significant at the $p < 0.05$ level after controlling for multiple comparisons using the Holm-Bonferroni procedure.

3.4.3 Logistic regression

Following Chen and Dredze (2018) and Broniatowski et al. (2016), we also conducted an analysis designed to test our model's predictions for whether a tweet was likely to be shared at least once treating this as a binary classification task. We once again conducted three rounds of model fitting, with each round containing two model fits.

1. *Explanation* a standard logistic regression model fit to all of the data using bidirectional stepwise elimination with AIC as the minimization criterion.
2. *Prediction* a logistic regression model with L_1-norm regularization fit to two-thirds of the data using threefold cross-validation, and evaluated against the remaining third. Here, we randomly undersampled tweets with no retweets to control for class imbalance, again comparing our model's results to "null" and "saturated" variants.

In each case, we used the same set of covariates as in the linear regression analyses, where the target variable was whether or not a given tweet had at least one retweet. Our second and third round of model fitting followed the same procedure used for the linear regression models, only substituting logistic regression for OLS regression.

Table 5 Best-fitting linear regression model predicting number of retweets per follower for tweets with at least one retweet

Covariate	β (*SE*)	*t*
User verified?	– 2.51 (0.01)	– 18.40***
Topic 2	0.63 (0.14)	4.49***
Topic 17	0.52 (0.14)	3.75***
Topic 2 × verbatim	– 0.06 (0.01)	– 5.01***
(Intercept)	– 2.61 (0.79)	– 3.32***

β Linear regression coefficient. *SE* standard error. Coefficients represent an increase in retweets per follower (measured in logits) for each unit increase in the dependent variable. All topic proportions are also measured in logits. User Verified? is a dummy variable indicating whether the account tweeting the message corresponded to a verified user (1) or not (0). Total model $R^2 = 0.23$

***p < 0.001

Table 6 Linear regression model performance compared to null and saturated models

	R^2		MSE	
	Training	Holdout	Training	Holdout
Null	0.00	− 0.01	3.92	4.25
User verification only	0.19	0.22	3.19	3.29
Round 1 (LASSO)	0.23	0.24	3.01	3.21
Round 2 (complex OLS)	0.24	0.25	2.97	3.16
Round 3 (simple OLS)	0.22	0.25	3.07	3.16
Saturated	0.68	− 1.38	1.25	5.81

R^2 the coefficient of determination, *MSE* mean squared error, *OLS* ordinary least squares. Round 1 = the model fit by LASSO regression. Round 2 = an OLS regression model retaining only variables that replicated across the LASSO and bidirectional elimination methods. Round 3 = an OLS regression model retaining only variables that were significant in the explanatory model after controlling for multiple comparisons using the Holm-Bonferroni procedure

Table 7 Best-fitting logistic regression model predicting whether a given tweet will be shared at least once

Covariate	β (*SE*)	*z*	OR
User verified?	2.38 (0.13)	17.95***	10.76
Contains media?	0.75 (0.09)	8.56***	2.12
Topic 28	0.45 (0.09)	8.56***	1.56
Topic 39	0.28 (0.09)	3.04**	1.32
Topic 12 × neutral	0.31 (0.10)	3.24**	1.37
Topic 12 × pro-vaccine	0.19 (0.10)	1.99*^{,NS}	1.21
Topic 12 × anti-vaccine	0.13 (0.10)	1.31	1.14

Coefficients represent a unit increase in the log-likelihood that a given message will be retweeted at least once for each unit increase in the dependent variable. All topic proportions are also measured in logits. User verified? and contains media? are dummy variables indicating whether the account tweeting the message corresponded to a verified user or contained media (1) or not (0), respectively

β Logistic regression coefficient, *SE* standard error, *OR* odds ratio, ^{NS} not significant after controlling for multiple comparisons using the Holm-Bonferroni procedure

***p < 0.001, **p < 0.01, *p < 0.05

4 Results

4.1 Linear regression analysis

Table 5 shows the explanatory linear regression model resulting from our selection procedure (intermediate models are presented in the Appendix).

Among the variables posited by our model, Topic 2 has the largest positive coefficient, indicating that messages containing the gist that vaccines cause autism are more likely to be shared. Topic 17, corresponding to vaccinations for pets, was also more likely to be shared. A negative interaction term between Topic 2 and verbatim features indicates that sharing for this topic decreases for longer tweets with more words. Finally, tweets from non-verified accounts were shared significantly more than tweets from verified accounts.

Beyond this explanatory analysis, we compared our model's performance to a Null model (only a constant predictor), a model containing only one feature (user verification), and a saturated model (containing all features and interactions) (Busemeyer and Wang 2000). Table 6 shows that the predictive power of the four factors shown in 5 improves upon simpler models, and equals or exceeds more complex models on holdout data.

4.2 Logistic regression analysis

Table 7 shows the explanatory logistic regression model resulting from our selection procedure (intermediate models are presented in the Appendix). Among the variables posited by our model, and consistent with prior work (Chen and Dredze 2018; Broniatowski et al. 2016), user verification and the presence of media are both significantly associated with more sharing. Additionally, topic 28 — corresponding to a "link" betweeen vaccines and autism—and topic 39—corresponding to a statistical description of vaccine adverse event rates—were both associated with more sharing. Finally, topic 12—associated with "big pharma" conspiracy theories—led to more sharing when the associated sentiment was neutral.

Table 8 shows that the predictive power of the factors shown in Table 7 improves upon simpler models, and exceeds more complex models on holdout data.

5 Discussion

Results of our analysis support FTT's implications. Topic 2, expressing a causal gist, is the strongest predictor of retweets per follower, replicating across multiple methodologies. Notably, this effect was attenuated when messages in this topic contained more difficult verbatim features, providing some evidence in favor of the role of multiple mental representations and the hierarchy of gist. Furthermore, consistent with the weaker role of verbatim representations, Topic 39, expressing verbatim statistics about vaccine-related adverse events, predicted only the likelihood of a single retweet.

We found support for several of our model's other parameters: the role of motivational factors on the first retweet is illustrated by the significant positive effects of user verification and vivid media on whether a message is retweeted at least once. Notably, user verification has a positive effect on the likelihood that a tweet is retweeted at least once, but a *negative* effect on the total number of retweets per follower, indicating that, without meaningful content, tweets from verified users are even less likely to go viral than tweets from unverified users. This may be because these accounts simply tend to generate more content and have more followers. Finally, Topic 28, which perhaps expresses similar thematic content as Topic 2

Table 8 Logistic regression model performance compared to null and saturated models

	Accuracy		Precision		Recall		F_1 score	
	Training	Holdout	Training	Holdout	Training	Holdout	Training	Holdout
Null	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50
User verification only	0.56	0.56	0.87	0.86	0.13	0.16	0.23	0.26
Round 1 (L_1-norm regularization)	0.64	0.63	0.58	0.65	0.67	0.56	0.62	0.60
Round 2 (complex logistic)	0.63	0.62	0.59	0.63	0.64	0.56	0.62	0.60
Round 3 (simple logistic)	0.64	0.64	0.59	0.67	0.65	0.56	0.62	0.61
Saturated	0.78	0.58	0.75	0.58	0.80	0.56	0.77	0.57

Round 1 = the model fit by L_1-norm regularization. Round 2 = a standard logistic regression model retaining only variables that replicated across the L_1-norm regularization and bidirectional elimination methods. Round 3 = an standard logistic regression model retaining only variables that were significant in the explanatory model after controlling for multiple comparisons using the Holm-Bonferroni procedure without expressing a gist—i.e., it mentions a "link' between vaccines and autism, but not a causal connection—only increased the likelihood of the first share.

Our results extend our recent findings on Facebook data (Broniatowski et al. 2016) where we showed that only gist was associated with increasing numbers of Facebook shares of vaccine-related news articles, whereas gist, verbatim statistics, and vivid media all predicted at least one share. Thus, this study extends our results from the most popular social media platform (as of 2018, 68% of the US adult population is on Facebook (Smith and Anderson 2018) to multiple social media.

5.1 Limitations and directions for future work

Our findings are limited by difficulties operationalizing the core constructs of FTT. Although LDA topics may be associated with gist in many cases, they do not in general capture the construct of meaning in context, which depends both on the prior knowledge of the observing subject and the stimulus. Future work should therefore focus on methods to extract gists given a candidate set of messages in the context of the knowledge likely to be widely held within a given online community. Similarly, representational weights are expected to vary with individual social media user accounts; therefore, proxy attributes of tweet content, such as emotion word dictionaries, readability, or verbatim features, will likely be noisy. Indeed, the results in Table 4 show that nominally distinct constructs such as valence, arousal, dominance were conflated. Similarly, discrete emotion states group into dimensions primarily reflecting overall sentiment. Future work may profitably focus on deriving relevant psychometric features given a sufficiently large set of tweets that might be used to characterize stable personality traits and other individual differences (e.g., Quercia et al. 2011; Golbeck et al. 2011). Importantly, vaccine sentiment evaluated on individual tweets may be a poor proxy for values stored in long-term memory. Indeed, those who support and oppose vaccination may agree on several relevant values, such as "saving lives is good" or "avoid harm", while disagreeing on the specific factors that might save lives or avoid harm. Specifically, those who oppose vaccination may contend that vaccines cause harm whereas those who support vaccination may be more concerned about harms caused by viral illnesses. It is therefore not surprising that the sentiment of vaccine messages was less of a predictive feature in our model.

Our results also highlight the complex role that emotion may play in both gist and motivation. Although the effects of emotional keywords did not replicate across multiple methods, Topic 17, corresponding to vaccinations for pets, significantly increased retweets per follower and likely has both motivational and gist components. Content referring to pets and other animals [e.g., cat videos and pandas (Hsee and Rottenstreich 2004; Myrick 2015)] tend to trigger emotional responses that can both increase arousal, a motivational factor, while also facilitating the retrieval of gist principles that influence decisions (Rivers et al. 2008), such as those to share online. Thus, future work should better characterize the role of strong emotion on both meaningful and motivational factors associated with online sharing.

6 Conclusions

In this paper, we propose a formal model of information sharing online based on FTT. Our model incorporates elements into its formulation that capture motivational factors as well as factors associated with the extraction of meaning from the article's content. These factors are predictive of online sharing, allowing us to make novel predictions.

Overall, our model provides strong support for the roles of multiple mental representations, but especially causal—i.e., meaningful—gist, combined with motivational factors. It appears that motivation aids a given tweet to be retweeted at least once; however, once retweeted, gist may be the engine underlying its virality.

Acknowledgement Preparation of this manuscript was supported in part by the National Institute of General Medical Sciences R01GM114771 to the first author.

Appendix

This Appendix contains details of our model fitting methodologies.

Linear regression

We first identified 1388 tweets that had been retweeted at least once. These tweets were used for subsequent analyses.

Bidirectional stepwise elimination

We implemented Bidirectional stepwise elimination using the R Project for Statistical Computing version 3.5.1. Specifically, we used the `step` function in R. The starting model contained main effects for the following features:

1. All 50 topic proportions (logit transformed)
2. Dummy variables for media and user verification
3. All three PCA dimensions for text complexity in Table 3
4. All three PCA dimensions for emotion in Table 4
5. tweet sentiment (positive, negative, or neutral)

Given this initial model, the `step` function successively added interaction terms (specifically between tweet sentiment, topic proportion, and text complexity) and removed other terms already in the model until AIC reached a local minimum. The resulting model is shown in Table 9.

Table 9 Linear regression model derived via bidirectional stepwise elimination minimizing AIC and explaining number of retweets per follower for tweets with at least one retweet

Covariate	β (*SE*)	*t*
User verified?	– 2.58 (0.14)	– 18.40***
Readability	– 7.83 (1.73)	– 4.52***
Topic 2	0.64 (0.14)	4.48***
Topic 5 × anti-vaccine	1.73 (0.54)	3.23**^{NS}
Topic 28 × anti-vaccine	0.95 (0.32)	3.02**^{NS}
Topic 30 × anti-vaccine	– 1.13 (0.39)	– 2.92**^{NS}
Topic 5 × pro-vaccine	2.19 (0.76)	2.90**^{NS}
Topic 5	– 1.26 (0.44)	– 2.85**^{NS}
Topic 30	0.59 (0.22)	2.70**^{NS}
Topic 16 × anti-vaccine	0.93 (0.36)	2.60**^{NS}
Negative emotions	– 0.13 (0.05)	– 2.58**^{NS}
Topic 27	0.41 (0.16)	2.56*^{NS}
Topic 10 × anti-vaccine	– 1.05 (0.41)	– 2.56*^{NS}
Topic 50 × readability	– 0.49 (0.19)	– 2.55*^{NS}
Topic 17 × verbatim	0.36 (0.14)	2.47*^{NS}
Topic 39	0.35 (0.14)	2.46*^{NS}
Topic 41	0.77 (0.31)	2.46*^{NS}
Topic 12 × pro-vaccine	– 1.48 (0.60)	– 2.46*^{NS}
Topic 14 × pro-vaccine	– 1.01 (0.43)	– 2.36*^{NS}
Topic 10	0.60 (0.26)	2.32*^{NS}
Topic 49 × anti-vaccine	– 0.98 (0.43)	– 2.30*^{NS}
Topic 45 × sentences	– 0.38 (0.17)	– 2.23*^{NS}
Topic 50 × sentences	0.23 (0.10)	2.22*^{NS}
Topic 17	0.34 (0.15)	2.20*^{NS}
Topic 28 × pro-vaccine	0.99 (0.46)	2.15*^{NS}
Topic 16	– 0.60 (0.29)	– 2.05*^{NS}
Topic 27 × readability	– 0.31 (0.15)	– 2.04*^{NS}
Topic 13	0.34 (0.17)	2.03*^{NS}
Topic 45	0.34 (0.17)	2.03*^{NS}
Topic 14	0.37 (0.19)	1.95
Topic 14 × anti-vaccine	– 0.78 (0.40)	– 1.94
Topic 2 × verbatim	– 0.26 (0.14)	– 1.89
Topic 19	0.27 (0.15)	1.76
Topic 31 × readability	– 0.30 (0.17)	– 1.75
Topic 11 × verbatim	0.33 (0.19)	1.75
Verbatim	1.89 (1.09)	1.73
Topic 28	– 0.43 (0.25)	– 1.68
Topic 29	0.25 (0.15)	1.66
Topic 30 × readability	– 0.27 (0.17)	– 1.63
Topic 24 × sentences	0.47 (0.29)	1.62
Topic 8 × readability	– 0.24 (0.15)	– 1.57
Topic 49	0.52 (0.33)	1.57

Table 9 (continued)

Covariate	β (*SE*)	*t*
ANEW	0.08 (0.05)	1.56
Topic 12 × readability	– 0.34 (0.23)	– 1.52
Topic 44	– 0.33 (0.22)	– 1.52
Topic 10 × pro-vaccine	– 0.74 (0.49)	– 1.51
Topic 8	0.23 (0.16)	1.46
Topic 32	0.24 (0.17)	1.45
Topic 33	0.21 (0.15)	1.41
Topic 24	0.28 (0.20)	1.41
Topic 30 × pro-vaccine	– 0.53 (0.45)	– 1.17
Topic 11	– 0.19 (0.16)	– 1.15
Topic 12 × anti-vaccine	– 0.43 (0.41)	– 1.05
Sentences	1.37 (1.40)	0.98
Topic 49 × pro-vaccine	– 0.47 (0.52)	– 0.92
Topic 31	– 0.14 (0.18)	– 0.81
Topic 16 × pro-vaccine	0.32 (0.43)	0.75
Anti-vaccine	– 2.96 (4.61)	– 0.64
Pro-vaccine	– 2.51 (5.38)	– 0.47
Topic 12	0.13 (0.34)	0.37
Topic 50	0.00 (0.17)	0.01
(Intercept)	8.35 (3.94)	2.12*^{NS}

β Linear regression coefficient. *SE* Standard error, ^{NS} not significant after controlling for multiple comparisons using the Holm-Bonferroni procedure. Coefficients represent an increase in retweets per follower (measured in logits) for each unit increase in the dependent variable. All topic proportions are also measured in logits. User Verified? is a dummy variable indicating whether the account tweeting the message corresponded to a verified user (1) or not (0). Total model $R^2 = 0.33$

***p < 0.001, **p < 0.01, *p < 0.05

LASSO regression

In parallel, we used LASSO regression to fit a model with the same predictors as above (including two- and three-way interactions between topic proportions, topic sentiment, and readability). We held out one third of our 1,388 datapoints, keeping the remaining data for training and test. LASSO regression was implemented using the `scikit-learn` package in Python 2.7. Specifically, we used the `LassoCV` function with 10,000 iterations and with `eps=1e-4`. The resulting model is shown in Table 10.

Model concordance

Although the bidirectional and LASSO models have several similarities, they also have many differences. We therefore fit a second pair of OLS models using only

Table 10 Linear model derived via LASSO regression and predicting number of retweets per follower for tweets with at least one retweet

Covariate	β
User verified?	– 2.14
Topic 2	0.15
Topic 2 × verbatim	– 0.09
Negative emotions	– 0.07
Topic 42 × pro-vaccine	– 0.07
Topic 17	0.06
ANEW	0.06
Topic 6 × verbatim × anti-vaccine	0.05
Topic 33 × sentences × anti-vaccine	– 0.03
Topic 17 × readability	0.02
Topic 1 × sentences	– 0.02
Topic 17 × readability × pro-vaccine	0.02
Topic 17 × verbatim × pro-vaccine	0.02
Topic 5 × verbatim × anti-vaccine	0.02
Topic 12 × anti-vaccine	– 0.01
Topic 16 × pro-vaccine	– 0.01
Topic 5 × sentences × pro-vaccine	0.01

β Linear regression coefficient. Coefficients represent an increase in retweets per follower (measured in logits) for each unit increase in the dependent variable. All topic proportions are also measured in logits. User verified? is a dummy variable indicating whether the account tweeting the message corresponded to a verified user (1) or not (0)

Table 11 OLS regression model using covariates that appeared in both LASSO and bidirectional models, and explaining number of retweets per follower for tweets with at least one retweet

Covariate	β (*SE*)	*t*
User verified?	– 2.61 (0.14)	– 18.52***
Topic 2 × verbatim	– 0.07 (0.01)	– 4.95***
Topic 2	0.60 (0.14)	4.29***
Topic 17	0.42 (0.14)	3.00**
Topic 12 × anti-vaccine	– 0.53 (0.20)	– 2.69**^{NS}
Negative emotions	– 0.13 (0.05)	– 2.59**^{NS}
Neutral × Topic 16	– 0.57 (0.25)	– 2.25*^{NS}
ANEW	0.08 (0.05)	1.70
Anti-vaccine × Topic 16	0.26 (0.18)	1.44
Topic 12 × neutral	0.31 (0.27)	1.14
Topic 12 × pro-vaccine	– 0.20 (0.32)	– 0.64
Pro-vaccine × Topic 16	– 0.15 (0.28)	– 0.55
(Intercept)	– 4.18 (1.25)	– 3.35***

β Linear regression coefficient. Coefficients represent an increase in retweets per follower (measured in logits) for each unit increase in the dependent variable. All topic proportions are also measured in logits. User verified? is a dummy variable indicating whether the account tweeting the message corresponded to a verified user (1) or not (0)

Table 12 Logistic regression model derived via bidirectional stepwise elimination minimizing AIC and explaining number of retweets per follower for tweets with at least one retweet

Covariate	β (*SE*)	*z*	OR
User verified?	2.43 (0.14)	17.72***	11.33
Contains media?	0.84 (0.09)	9.06***	2.32
Readability	5.04 (1.25)	4.03***	154.96
Topic 28	0.41 (0.10)	4.00***	1.50
Topic 39 × verbatim	0.46 (0.13)	3.54***	1.59
Topic 38	0.37 (0.11)	3.38***^{NS}	1.44
Topic 50	0.36 (0.11)	3.22**^{NS}	1.43
Topic 29	0.33 (0.11)	3.12**^{NS}	1.39
Topic 19 × sentences	– 0.39 (0.13)	– 3.08**^{NS}	0.68
Verbatim × anti-vaccine	0.24 (0.08)	3.06**^{NS}	1.27
Topic 28 × readability	0.33 (0.11)	3.04**^{NS}	1.38
Topic 39 × sentences	0.28 (0.09)	2.95**^{NS}	1.32
Topic 50 × verbatim	– 0.30 (0.10)	– 2.93**^{NS}	0.74
Topic 8 × readability	0.34 (0.12)	2.91**^{NS}	1.40
Topic 36 × pro-vaccine	0.91 (0.31)	2.91**^{NS}	2.49
Topic 15 × sentences	0.25 (0.09)	2.69**^{NS}	1.28
Topic 25	0.40 (0.16)	2.52*^{NS}	1.49
Topic 39 × pro-vaccine	– 1.18 (0.47)	– 2.51*^{NS}	0.31
Topic 23	– 0.31 (0.13)	– 2.42*^{NS}	0.74
Topic 3	0.28 (0.12)	2.32*^{NS}	1.32
Topic 14 × verbatim	– 0.26 (0.12)	– 2.25*^{NS}	0.77
Topic 25 × anti-vaccine	– 0.53 (0.24)	– 2.23*^{NS}	0.59
Topic 20 × sentences	– 0.33 (0.15)	– 2.21*^{NS}	0.72
Topic 8 × verbatim	0.26 (0.12)	2.11*^{NS}	1.30
Topic 39	0.31 (0.15)	2.08*^{NS}	1.36
Topic 33 × verbatim	0.25 (0.12)	2.08*^{NS}	1.29
Topic 48 × sentences	– 0.24 (0.11)	−2.06*^{NS}	0.79
Topic 16 × anti-vaccine	0.50 (0.25)	2.03*^{NS}	1.65
Topic 19	0.21 (0.11)	1.96*^{NS}	1.23
Topic 17	0.21 (0.11)	1.93	1.24
Topic 4 × verbatim	0.23 (0.12)	1.92	1.25
Sentences	– 2.84 (1.48)	– 1.92	0.06
Topic 4 × sentences	– 0.21 (0.11)	– 1.90	0.81
Topic 43	0.20 (0.11)	1.89	1.22
Topic 47 × sentences	– 0.29 (0.15)	– 1.87	0.75
Topic 12	0.43 (0.23)	1.86	1.53
Topic 14	0.20 (0.11)	1.84	1.22
Topic 10	– 0.22 (0.12)	– 1.81	0.80
Topic 35 × readability	0.23 (0.13)	1.78	1.26
Topic 4 × readability	0.21 (0.12)	1.76	1.24
Topic 25 × sentences	0.18 (0.10)	1.76	1.20
Topic 21	0.21 (0.12)	1.74	1.24

Table 12 (continued)

Covariate	*β* (*SE*)	*z*	OR
Topic 31	0.20 (0.12)	1.69	1.22
Topic 36	− 0.31 (0.18)	− 1.66	0.74
Topic 27	0.18 (0.11)	1.61	1.20
Topic 4	0.18 (0.11)	1.60	1.20
Topic 18 × sentences	− 0.18 (0.12)	− 1.56	0.83
Topic 6 × sentences	0.18 (0.12)	1.56	1.20
Negative emotions	− 0.05 (0.04)	− 1.55	0.95
Topic 27 × readability	0.16 (0.10)	1.54	1.17
Topic 43 × verbatim	− 0.19 (0.12)	− 1.53	0.83
Topic 35 × verbatim	− 0.19 (0.13)	− 1.52	0.82
Topic 15 × readability	− 0.19 (0.13)	− 1.49	0.82
Topic 17 × verbatim	0.16 (0.11)	1.43	1.17
Topic 12 × pro-vaccine	0.61 (0.43)	1.41	1.84
Topic 18 × readability	0.19 (0.13)	1.41	1.20
Topic 18	0.16 (0.12)	1.38	1.17
Topic 5	− 0.21 (0.15)	− 1.38	0.81
Topic 6	0.16 (0.12)	1.37	1.17
Topic 35	0.17 (0.12)	1.35	1.18
Topic 47	− 0.18 (0.14)	− 1.30	0.83
Verbatim	1.82 (1.41)	1.29	6.15
Topic 12 × anti-vaccine	− 0.33 (0.28)	− 1.18	0.72
Topic 36 × anti-vaccine	0.28 (0.27)	1.05	1.32
Topic 25 × pro-vaccine	− 0.24 (0.24)	− 1.00	0.78
Topic 39 × anti-vaccine	− 0.17 (0.24)	− 0.69	0.85
Topic 8	0.09 (0.13)	0.67	1.09
Topic 16	− 0.11 (0.20)	− 0.57	0.89
Topic 16 × pro-vaccine	− 0.15 (0.29)	− 0.50	0.86
Verbatim × pro-vaccine	0.03 (0.09)	0.38	1.04
Topic 20	− 0.05 (0.12)	− 0.38	0.96
Topic 48	0.04 (0.11)	0.37	1.04
Topic 15	0.04 (0.12)	0.35	1.04
Topic 33	0.03 (0.13)	0.24	1.03
Anti-vaccine	− 0.32 (2.32)	− 0.14	0.73
Pro-vaccine	0.16 (3.13)	0.05	1.17
(Intercept)	12.82 (2.67)	4.80***	

Coefficients represent a unit increase in the log-likelihood that a given message will be retweeted at least once for each unit increase in the dependent variable. All topic proportions are also measured in logits. User verified? and Contains Media? are dummy variables indicating whether the account tweeting the message corresponded to a verified user or contained media (1) or not (0), respectively

β Logistic regression coefficient, *SE* standard error, *OR* odds ratio, [NS] not significant after controlling for multiple comparisons using the Holm-Bonferroni procedure

***p < 0.001, **p < 0.01, *p < 0.05

Table 13 Logistic regression model derived via L_1-norm regularization and predicting number of retweets per follower for tweets with at least one retweet

Covariate	β	OR
User verified?	1.06	2.89
Contains media?	0.57	1.76
Topic 12 × anti-vaccine	– 0.13	0.88
Topic 3 × pro-vaccine	– 0.09	0.92
Topic 39	0.08	1.08
Negative emotions	– 0.06	0.94
Topic 25 × verbatim × anti-vaccine	– 0.06	0.94
Topic 2 × verbatim	– 0.05	0.96
Topic 3 × sentences × pro-vaccine	0.04	1.04
Topic 4 × sentences	– 0.03	0.97
Topic 17 × readability × pro-vaccine	0.03	1.03
Topic 3 × verbatim × pro-vaccine	– 0.03	0.97
Topic 37 × readability	– 0.03	0.97
Topic 49 × anti-vaccine	– 0.02	0.98
Topic 28	0.02	1.02
Topic 19 × sentences	– 0.01	0.99
Topic 42 × sentences × anti-vaccine	0.00	1.00

β Logistic regression coefficient, *SE* standard error, *OR* odds ratio

Coefficients represent a unit increase in the log-likelihood that a given message will be retweeted at least once for each unit increase in the dependent variable. All topic proportions are also measured in logits. User Verified? and Contains Media? are dummy variables indicating whether the account tweeting the message corresponded to a verified user or contained media (1) or not (0), respectively

those covariates that appeared in both the bidirectional and LASSO models (see Table 11). Finally, we removed all covariates that were not significant after multiple comparisons using the Holm-Bonferroni procedure to generate our final model, shown in Table 5.

Logistic regression

We next compared the 1388 tweets that had been retweeted at least once to the remaining tweets in our sample—i.e., those that had not been retweeted.

Bidirectional stepwise elimination

We once again implemented Bidirectional stepwise elimination using the `step` function in R. The starting model contained the same main effects and scope as in the linear regression model. The resulting model is shown in Table 12.

Table 14 Standard logistic regression model using covariates that appeared in both L_1-norm and bidirectional models, and explaining likelihood of at least one retweet

Covariate	β	OR	
User verified?	2.39 (0.13)	18.02***	10.90
Contains media?	0.73 (0.09)	8.30***	2.08
Topic 28	0.45 (0.09)	4.74***	1.56
Topic 12 × neutral	0.31 (0.10)	3.14**	1.36
Topic 39	0.27 (0.09)	2.90**	1.31
Topic 12 × pro-vaccine	0.19 (0.10)	1.96	1.21
Topic 12 × anti-vaccine	0.12 (0.10)	1.25	1.13
Topic 4 × sentences	– 0.05 (0.07)	– 0.71	0.95
Negative emotions	0.01 (0.03)	0.32	1.01
Topic 19 × sentences	0.02 (0.07)	0.26	1.02
(Intercept)	2.01 (0.67)	3.01**	7.44

β Logistic regression coefficient, *SE* standard error, *OR* odds ratio

Coefficients represent a unit increase in the log-likelihood that a given message will be retweeted at least once for each unit increase in the dependent variable. All topic proportions are also measured in logits. User Verified? and Contains Media? are dummy variables indicating whether the account tweeting the message corresponded to a verified user or contained media (1) or not (0), respectively

***p < 0.001, **p < 0.01

L_1-norm regularization

In parallel, we used L_1-norm regularization to fit a model with the same predictors as above (including two- and three-way interactions between topic proportions, topic sentiment, and readability). We retained the 1,388 datapoints that had been retweeted at least once and randomly sampled another 1,388 datapoints from the remaining data to account for class imbalance. We held out one third of this combined dataset, keeping the remaining data for training and test. Logistic regression with L_1-norm regularization was implemented using the `scikit-learn` package in Python 2.7. Specifically, we used the `LogisticRegressionCV` function with 4,000 iterations and the `liblinear` solver. The resulting model is shown in Table 13.

Model concordance

We once again resolved differences between the two models by fitting a second pair of logistic regression models using only those covariates that appeared in both the bidirectional and L_1-norm models (see Table 14). Finally, we removed all covariates that were not significant after multiple comparisons using the Holm-Bonferroni procedure to generate our final model, shown in Table 7.

References

Akaike H (1976) Canonical correlation analysis of time series and the use of an information criterion. In: Mathematics in science and engineering, vol 126, Elsevier, pp 27–96
Bakshy E, Messing S, Adamic LA (2015) Exposure to ideologically diverse news and opinion on facebook. Science 348(6239):1130–1132
Bansal S (2018) textstat:memo: python package to calculate readability statistics of a text object—paragraphs, sentences, articles.https://github.com/shivam5992/textstat. Accessed 16 June 2014
Berger J, Milkman KL (2012) What makes online content viral? J Mark Res 49(2):192–205
Betsch C, Brewer NT, Brocard P, Davies P, Gaissmaier W, Haase N, Leask J, Renkewitz F, Renner B, Reyna VF et al (2012) Opportunities and challenges of Web 2.0 for vaccination decisions. Vaccine 30(25):3727–3733
Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3:993–1022
Bradley MM, Lang PJ (1999) Affective norms for english words (anew). The NIMH Center for the Study of Emotion and Attention, University of Florida, Gainesville
Brewer NT, Chapman GB, Rothman AJ, Leask J, Kempe A (2017) Increasing vaccination: putting psychological science into action. Psychol Sci Public Interest 18(3):149–207
Broniatowski D, Reyna V (2018) A formal model of fuzzy-trace theory. Decision 5(4):205–252
Broniatowski DA, Klein EY, Reyna VF (2015) Germs are germs, and why not take a risk? Patients' expectations for prescribing antibiotics in an Inner-City Emergency Department. Med Decis Mak 35(1):60–67
Broniatowski DA, Hilyard KM, Dredze M (2016) Effective vaccine communication during the disneyland measles outbreak. Vaccine 34(28):3225–3228 7
Broniatowski DA, Jamison AM, Qi S, AlKulaib L, Chen T, Benton A, Quinn SC, Dredze M (2018) Weaponized health communication: twitter bots and Russian trolls amplify the vaccine debate. Am J Public Health 108(10):1378–1384
Busemeyer JR, Wang YM (2000) Model comparisons and model selections based on generalization criterion methodology. J Math Psychol 44(1):171–189
Buttenheim AM, Asch DA (2016) Leveraging behavioral insights to promote vaccine acceptance: one year after disneyland. JAMA Pediatr 170(7):635–636
Cacioppo JT, Petty RE, Feng Kao C (1984) The efficient assessment of need for cognition. J Personal Assess 48(3):306–307
Cacioppo JT, Feinstein JA, Jarvis WBG (1996) Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. Psychol Bull 119(2):197
Centola D (2010) The spread of behavior in an online social network experiment. Science 329(5996):1194–1197
Centola D (2011) An experimental study of homophily in the adoption of health behavior. Science 334(6060):1269–1272
Chang J, Gerrish S, Wang C, Boyd-Graber JL, Blei DM (2009) Reading tea leaves: how humans interpret topic models. In: Advances in neural information processing systems, pp 288–296
Chen T, Dredze M (2018) Vaccine images on twitter: analysis of what images are shared. J Med Internet Res 20(4):e130
Cheng J, Adamic L, Dow PA, Kleinberg JM, Leskovec J (2014) Can cascades be predicted? In: Proceedings of the 23rd international conference on World wide web, ACM, pp 925–936
Chou WYS, Oh A, Klein WM (2018) Addressing health-related misinformation on social media. Jama 320(23):2417–2418
Cokely ET, Kelley CM (2009) Cognitive abilities and superior decision making under risk: a protocol analysis and process model evaluation. Judgm Decis Mak 4(1):20–33
Cowling BJ, Fang VJ, Nishiura H, Chan KH, Ng S, Ip DK, Chiu SS, Leung GM, Peiris JM (2012) Increased risk of noninfluenza respiratory virus infections associated with receipt of inactivated influenza vaccine. Clin Infect Dis 54(12):1778–1783
Curseu PL (2006) Need for cognition and rationality in decision-making. Stud Psychol 48(2):141
Diehl JJ, Bennetto L, Young EC (2006) Story recall and narrative coherence of high-functioning children with autism spectrum disorders. J Abnorm Child Psychol 34(1):83–98
Dredze M, Broniatowski DA, Hilyard KM (2016a) Zika vaccine misconceptions: a social media analysis. Vaccine 34(30):3441–3442

Dredze M, Broniatowski DA, Smith MC, Hilyard KM (2016b) Understanding vaccine refusal: why we need social media now. AmJ Prev Med 50(4):550–552
Dredze M, Wood-Doughty Z, Quinn SC, Broniatowski DA (2017) Vaccine opponents' use of twitter during the 2016 us presidential election: implications for practice and policy. Vaccine 35(36):4670–4672
Frederick S (2005) Cognitive reflection and decision making. J Econ Perspect 19(4):25–42
Fukukura J, Ferguson MJ, Fujita K (2013) Psychological distance can improve decision making under information overload via gist memory. J Exp Psychol 142(3):658
Galván A (2017) The neuroscience of adolescence, 1st edn. Cambridge University Press, Cambridge, New York
Gernsbacher MA, Varner KR, Faust ME (1990) Investigating differences in general comprehension skill. J Exp Psychol 16(3):430
Gernsbacher MA (1996) The structure-building framework: what it is, what it might also be, and why. In: Britton BK, Graesser AC (eds) Models of understanding text. Psychology Press, New York, NY, pp 289–311
Golbeck J, Robles C, Edmondson M, Turner K (2011) Predicting personality from twitter. In: 2011 IEEE Third International conference on privacy, security, risk and trust (PASSAT) and 2011 IEEE Third International conference on social computing (SocialCom), IEEE, pp 149–156
Goldman SR, McCarthy KS, Burkett C (2015) Interpretive inferences in literature. In: Inferences during reading, p 386
Granovetter M, Soong R (1983) Threshold models of diffusion and collective behavior. J Math Sociol 9(3):165–179
Griffiths TL, Steyvers M, Tenenbaum JB (2007) Topics in semantic representation. Psychol Rev 114(2):211
Grinberg N, Joseph K, Friedland L, Swire-Thompson B, Lazer D (2019) Fake news on twitter during the 2016 us presidential election. Science 363(6425):374–378
Hsee CK, Rottenstreich Y (2004) Music, pandas, and muggers: on the affective psychology of value. J Exp Psychol 133(1):23
Jamison AM, Broniatowski D, Quinn SC (2019) Malicious actors on twitter: a guide for public health researchers. Am J Public Health 109:688–692
Kintsch W (1974) The representation of meaning in memory. Lawrence Erlbnum Associates, Hillsdale
Klein EY, Martinez EM, May L, Saheed M, Reyna V, Broniatowski DA (2017) Categorical risk perception drives variability in antibiotic prescribing in the Emergency Department: a mixed methods observational study. J Gen Intern Med 32(10):1083–1089
LaTour KA, LaTour MS, Brainerd C (2014) Fuzzy trace theory and "smart" false memories: implications for advertising. J Advert 43(1):3–17
LeBoeuf RA, Shafir E (2003) Deep thoughts and shallow frames: on the susceptibility to framing effects. J Behav Decis Mak 16(2):77–92
Liberali JM, Reyna VF, Furlan S, Stein LM, Pardo ST (2012) Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. J Behav Decis Mak 25(4):361–381
Linderholm T, Everson MG, van den Broek P, Mischinski M, Crittenden A, Samuels J (2000) Effects of causal text revisions on more- and less-skilled readers' comprehension of easy and difficult texts. Cogn Instr 18(4):525–556
Mandler JM (1983) What a story is. Behav Brain Sci 6(04):603–604
Mohammad SM, Kiritchenko S (2015) Using hashtags to capture fine emotion categories from tweets. Comput Intell 31(2):301–326
Mohammad SM, Turney PD (2013) Crowdsourcing a word-emotion association lexicon. Comput Intell 29(3):436–465
Myrick JG (2015) Emotion regulation, procrastination, and watching cat videos online: who watches internet cats, why, and to what effect? Comput Hum Behav 52:168–176
Mnøsted B, Sapieżyński P, Ferrara E, Lehmann S (2017) Evidence of complex contagion of information in social media: an experiment using twitter bots. PLoS ONE 12(9):e0184148
Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830
Pennington N, Hastie R (1991) A cognitive theory of juror decision making: the story model. Cardozo L Rev 13:519

Pennycook G, Rand DG (2018) Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. Cognition 188:39–50
Pennycook G, Cheyne JA, Koehler DJ, Fugelsang JA (2016) Is the cognitive reflection test a measure of both reflection and intuition? Behav Res Methods 48(1):341–348
Pennycook G, Cannon TD, Rand DG (2018) Prior exposure increases perceived accuracy of fake news. J Exp Psychol 147:1865–1880
Perrin A (2015) Social media usage: 2005–2015
Peters E, Levin IP (2008) Dissecting the risky-choice framing effect: numeracy as an individual-difference factor in weighting risky and riskless options. Judgm Decis Mak 3(6):435–448
Peters E, Västfjäll D, Slovic P, Mertz CK, Mazzocco K, Dickert S (2006) Numeracy and decision making. Psychol Sci 17(5):407–413
Petrovic S, Osborne M, Lavrenko V (2011) RT to win! Predicting message propagation in twitter. ICWSM 11:586–589
Plutchik R (2001) The nature of emotions: human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. Am Sci 89(4):344–350
Quercia D, Kosinski M, Stillwell D, Crowcroft J (2011) Our twitter profiles, our selves: predicting personality with twitter. In: 2011 IEEE Third International conference on privacy, security, risk and trust (PASSAT) and 2011 IEEE Third International conference on social computing (SocialCom), IEEE, pp 180–185
Quinn SC, Parmer J, Freimuth VS, Hilyard KM, Musa D, Kim KH (2013) Exploring communication, trust in government, and vaccination intention later in the 2009 H1n1 pandemic: results of a national survey. Biosecur Bioterror 11(2):96–106
Rapp DN, Pvd Broek, McMaster KL, Kendeou P, Espin CA (2007) Higher-order comprehension processes in struggling readers: a perspective for research and intervention. Sci Stud Read 11(4):289–312
Reese E, Haden CA, Baker-Ward L, Bauer P, Fivush R, Ornstein PA (2011) Coherence of personal narratives across the lifespan: a multidimensional model and coding method. J Cogn Dev 12(4):424–462
Reyna VF (2012) Risk perception and communication in vaccination decisions: a fuzzy-trace theory approach. Vaccine 30(25):3790–3797
Reyna VF, Adam MB (2003) Fuzzy-trace theory, risk communication, and product labeling in sexually transmitted diseases. Risk Anal 23(2):325–342
Reyna VF, Brainerd CJ (2008) Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. Learn Individ Differ 18(1):89–107
Reyna VF, Lloyd FJ (2006) Physician decision making and cardiac risk: effects of knowledge, risk perception, risk tolerance, and fuzzy processing. J Exp Psychol 12(3):179
Reyna VF, Estrada SM, DeMarinis JA, Myers RM, Stanisz JM, Mills BA (2011) Neurobiological and memory models of risky decision making in adolescents versus young adults. J Exp Psychol 37(5):1125
Reyna VF, Corbin JC, Weldon RB, Brainerd CJ (2016) How fuzzy-trace theory predicts true and false memories for words, sentences, and narratives. J Appl Res Mem Cogn 5(1):1–9
Riddell A (2014) Lda: 0.3.2. 10.5281/zenodo.592664. https://zenodo.org/record/592664. Accessed 16 July 2018
Rivers SE, Reyna VF, Mills B (2008) Risk taking under the influence: a fuzzy-trace theory of emotion in adolescence. Dev Rev 28(1):107–144
Rogers EM (2010) Diffusion of innovations. Simon and Schuster, New York
Romero DM, Meeder B, Kleinberg J (2011) Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In: Proceedings of the 20th international conference on World wide web, ACM, pp 695–704
Schley DR, Peters E (2014) Assessing economic value symbolic-number mappings predict risky and riskless valuations. Psychol Sci 25:753–761
Shmueli G et al (2010) To explain or to predict? Stat Sci 25(3):289–310
Silverman C (2016) This analysis shows how viral fake election news stories outperformed real news on facebook. Retrieved February 15, 2017, from https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook
Simon AF, Fagley NS, Halleran JG (2004) Decision framing: moderating effects of individual differences and cognitive processing. J Behav Decis Mak 17(2):77–93
Smith A, Anderson M (2018) Social media use in 2018. Pew Research Center 1
Stevens SS et al (1946) On the theory of scales of measurement. Science 103:677–680

Subrahmanian V, Azaria A, Durst S, Kagan V, Galstyan A, Lerman K, Zhu L, Ferrara E, Flammini A, Menczer F (2016) The darpa twitter bot challenge. Computer 49(6):38–46

Sundaram ME, McClure DL, VanWormer JJ, Friedrich TC, Meece JK, Belongia EA (2013) Influenza vaccination is not associated with detection of noninfluenza respiratory viruses in seasonal studies of influenza vaccine effectiveness. Clin Infect Dis 57(6):789–793

Swire B, Berinsky AJ, Lewandowsky S, Ecker UK (2017) Processing political misinformation: comprehending the trump phenomenon. R Soc Open Sci 4(3):160802

Trabasso T, Sperry LL (1985) Causal relatedness and importance of story events. J Mem Lang 24(5):595–611

Trabasso T, Secco T, Van Den Broek P (1984) Causal cohesion and story coherence. In: Mandl H, Stein NL, Trabasso T (eds) Learning and comprehension of text. Lawrence Erlbaum Associates, Hillsdale, NJ, pp 83–110

Trope Y, Liberman N (2010) Construal-level theory of psychological distance. Psychol Rev 117(2):440

Tsur O, Rappoport A (2012) What's in a hashtag?: content based prediction of the spread of ideas in microblogging communities. In: Proceedings of the fifth ACM international conference on Web search and data mining, ACM, pp 643–652

Tversky A, Kahneman D (1981) The framing of decisions and the psychology of choice. Science 211(4481):453–458

Van den Broek P (2010) Using texts in science education: cognitive processes and knowledge representation. Science 328(5977):453–456

van den Broek P, Helder A (2017) Cognitive processes in discourse comprehension: passive processes, reader-initiated processes, and evolving mental representations. Discourse Process 54:1–13

Vazquez MA (2016) Informe de Médicos de Pueblos Fumigados sobre Dengue-Zika y fumigaciones con venenos químicoshttp://alimentoyconciencia.com/informe-de-medicos-de-pueblos-fumigados-sobre-dengue-zika-y-fumigaciones-con-venenos-quimicos/. Accessed 06 Feb 2017

Vosoughi S, Roy D, Aral S (2018) The spread of true and false news online. Science 359(6380):1146–1151

Dr. David A. Broniatowski is Director of the Decision Making and Systems Architecture Laboratory, conducts research in decision making under risk, group decision making, system architecture, and behavioral epidemiology. This research program draws upon a wide range of techniques including formal mathematical modeling, experimental design, automated text analysis and natural language processing, social and technical network analysis, and big data. Current projects include a text network analysis of transcripts from the US Food and Drug Administration's Circulatory Systems Advisory Panel meetings, a mathematical formalization of Fuzzy Trace Theory—a leading theory of decision-making under risk, derivation of metrics for flexibility and controllability for complex engineered socio-technical systems, and using social media data to study fake news and why it spreads.

Dr. Valerie F. Reyna is Director of the Human Neuroscience Institute, Co-Director of the Cornell University Magnetic Resonance Imaging Facility, Co-Director of the Center for Behavioral Economics and Decision Research, and Professor of Human Development, Psychology, Cognitive Science, and Neuroscience (IMAGINE Program) at Cornell University. She is a leader in using memory principles and mathematical models to explain judgment and decision making, and helped initiate what is now a burgeoning area of research on developmental differences in judgment and decision making. She is a developer of fuzzy-trace theory, an influential model of the relation between mental representations and decision making that has been widely applied in law, medicine, and public health. Her recent work has focused on behavior change; neuroeconomics; rationality and risky decision making; and neuroscience models of decision making. She has applied fuzzy-trace theory to risk perception, numeracy, and medical decision making by both physicians and patients.